

SUBSTITUTE SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: GENSET SA
- (B) STREET: 24 RUE ROYALE
- (C) CITY: PARIS
- (E) COUNTRY: FRANCE
- (F) POSTAL CODE: 75008

(ii) TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC COMPOSITION
CONTAINING SAME AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
- (B) STREET: 2421 N.W. 41st Street, Suite A-1
- (C) CITY: Gainesville
- (D) STATE: Florida
- (E) COUNTRY: USA
- (F) ZIP: 32606

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER (unassigned)
- (B) FILING DATE: OCTOBER 18, 2001

(vii) PRIORITY APPLICATION DATA:

- (A) APPLICATION NUMBER 09/486,580
- (B) FILING DATE: FEBRUARY 25, 2000

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Frank C. Eisenschenk, Ph.D.
- (B) REGISTRATION NUMBER: 45,332
- (C) REFERENCE/DOCKET NUMBER: GEN-100D1

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4415 BASE PAIRS
- (B) TYPE: NUCLEOTIDE
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: Exon 1
- (B) LOCATION: 1836..1874

(ix) FEATURE:
 (A) NAME/KEY: Exon 2
 (B) LOCATION: 3394..3577

(ix) FEATURE:
 (A) NAME/KEY: Exon 3
 (B) LOCATION: 4161..4380

(ix) FEATURE:
 (A) NAME/KEY: start CDS
 (B) LOCATION: 3406..3408

(ix) FEATURE:
 (A) NAME/KEY: stop CDS
 (B) LOCATION: 4276..4278

(ix) FEATURE:
 (A) NAME/KEY: polyadenylation site
 (B) LOCATION: 4374..4379

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACACCATTG TCTTCATGTA ACCCCATTAG CTATACCCTC TAGTGCAAGG AAACCATAGG	60
GCCTAGGTCA CACCATGAGG CTGCNCTTAC AAGTTATGCA AAAACTATGG ACTTGGGAGA	120
CCTGTGCGTA ACAACATCAC ACNCCAAATT TAACCAGCTC TCCCCATAAC AGCACGCTCA	180
TGTGTTACTG AGGAAATGCC TGTGGATTGG AGTGTGTTCT GTGTGCAGGA GGCTGGTCCA	240
GGTTTCACTT CTGCAGGACA CTGGACGTTT CCCAAAACCA GCAGACTTTC CCCACGTGCA	300
CACACACCCC TTCTCATTTC GCCTCTACAT CCATATCCAC TGGGCCCTTC AGGCACCTAC	360
TAATGCCCTA GAACCTAAAA CCATCATCTG GGGCCAGTT CCCTGAATGG CCCTAATCTC	420
TTCCTCTGCT GGAATGAGTC CAGTGCCAC TTCCTCCAAC GGTGAAATTG CTGGGCTGCT	480
ACAGATCAGG AACTCACTGC TTCCTCATAG GGGCAGCCGA CTTCACTGCT CTGCAACAGC	540
GACCACCCCT AGCGAGGCTT GAGATGCCTC TTGCCTCCTT AAGACTGAGG GAGACGCTTC	600
AGCTCTCACT CCACTGCCCA AAGTCCTCCA CAGCGCGGTG CCTGCTGCCT TCACACAGAG	660
CTGCAGGGGN AGGTCTGTG TATCCGGCCT GCTGGACCAG CGCTGTGCAC AACCCCTCCA	720
TGGCAACAGT GGCTGCCCGG CCTGCACACT GGGCTTGGCA ACCTCGCTGT AGGTATTAT	780
TCCCTCAGGA GTGACTGCAT TCTTTTCCA TTTCCAGAAA ACTGATGCCA TTTACCTCAC	840
TATGAGGAGG AGGAGGAGGA GGAGGGTGGA GAGTGGTACA TTTAAAATG TGCACTATT	900
TCCCTAGGAC TCCCCCTCAA ATAACCCAGG AGGGACCATA CCAGCTCATT CCTGTGTATC	960
CCAAGCATAN GAGTAATCAT CCCACTCATG CTGAGTGTAT GGTGGCCATT AAGCCTGCC	1020
TGAACGGCT TTAGAACAAAG GTGTTGAGC ACACAGCACC GTCTTGCTGC CACCTGGCC	1080
CCCTCCCTTG TGAGACCTCT GAGACACATT NAGGTCTCAC CTAAAAATCT CAGGATTCT	1140

AGGCCAAAN CGGTCTAAA AAATTGTTCA GTCTGAACTC TCTAAGGTCA AGAGAAGAGG	1200
TGGTTGCTCC CTCTAAGAAA CCACATGTTG CATGTACATC CTTAATTCCG GAAAGTCAA	1260
CAAACCTGCC CTGCTTAGCA ACACAAGCCG AGGTGGTA CTCCTCACCC GGGCATCTC	1320
CAACACACCT GTTTGTCAA ACAGCTTG A TTTGTTTTA TAGTTGGACC CCAGGTTCCC	1380
AGGAGGCTGG TTCAGGCCAT ATTCCAAATC CTCATCTGTG TGTGAGTGGC ATTCTTAGCC	1440
TAGCCTCCTT ACAGGGTGG A TACTATGATA CACAGCCAGG CTGCTCCAGT GGCTTCAAT	1500
ATTCTTTGG TCCAGATAGT TCAGCCTCAG CACCAGTGTA GGCATCACAG GGTCAATTGT	1560
CTTAGGAGTC ATGGAGAATT CATAGTTGGT AGCTACCTGG GCCTGGCCAG GGCTGACCAG	1620
AGACAAGGCA TCCCTCTGTG AACTCCTATT TTAATGCCAG CTTCCAACA AATTCTCAA	1680
CTGCTCTTAC CAGCAGGTAT TTAAACTACT CAATAGAAAG TAACCCTGAA AATTAGGACA	1740
CCTGTTCCA AAAGACCCTT AAATAGGGGA AGTCCTTCN CTGCTTGTGC ACAGCTGCTG	1800
ATGTGGCAAC ATGAGGCCTG GGACAGGGGA CTGCTCTG CCCACTCTGG TAGCCTCACG	1860
TAGCTTAACA ATCTGTCAGT AATACAATAC AAAACTTAAA CTTTCATACT GCGGTTCCAC	1920
CCAGGAAGCT GTGTTCCCAA TCTGACCGT GATTATGGGG CCACCTCAGA GGGNACCCAG	1980
TGAGGGATA TTTTGCCATC TGGGACTGTT GGTTGCTGGG GGCAGTGGCT ATGAGCTCAG	2040
TTAATAAACT CAAGCAGTTT CCTTCCAAAC ACACATGTCC TACTTAACGT GTCCAACAGA	2100
GATGATCATA CTCATANGCT GCTAAAACAT TANTTTTATT TTGAGAAAAG TCTATTGATG	2160
TTCTGGCCC ATGGAGTTT CATTNATTA NTTTATTTAT TTTGCAGAGA TGGAGTCTCA	2220
CTATGTTGCT CAAGCTGGTC TCCAACCTCCT GGGCTCAAGC GATCTCCTA CTTTGGCCTT	2280
TGAAAGCGCT GAGATTGCC TGTGAGCCA TCATGGGGC TCACTGGCCC ACTGATTAAT	2340
CAGATTAATT GTTTTTGCT ATTGAANTTG TTTGACTTCC TTGTATATTC GGATATTTAC	2400
CCATTCTAAC ACGTAGGGTT TGCAAATATT TTCTCTCATG TTCTGTGTTG CCTTTCACT	2460
CAGTTGATGG TTTCTTTGC TGTGCAAGGTG CTTTAGTGT CAACGCAGCC CCGCTTGTCT	2520
ATTTCCATT TTATTGCCTG TCCCTTGAT GTCATAGCCA AGAAATAATT GCCCAGATTA	2580
ATGTCAAAAA GCTTTATCCC TATATATTCT TCTAGTAGTT TATGGTTCA GATCTTATGT	2640
TTAGGTCTTC AATCCATTGA GTTGATTTT GTATGTGGTA TAAGAAAAAA GACCACATGT	2700
ATACATATCT CAAATTCTAA GGTAGTATAT ATTAGACACA TACAATGTGT CTATTTACAC	2760
ACATTGAGCT GAAAATAATA AACATATTAA TATCTTCAA TCAACTCTAT CTCTATCTCA	2820
CTGAACCTGT TTCACCTATA GCCTGATGAG GTTGCTGTCC TCTCTACCC AGCTCCTATA	2880
GGAGACTGCT CATCCCCTAA CCTCAAAAAC CCCTTCATGA GGGTGATAAT GCCCTTGAAT	2940
CCTGCAATGA ATTAGTTCTC TACTACAGTG GAATTCAAGGT CTGTTATGAG GGTCTGGATC	3000

TCTGAAGAGA AGAGCTCTCA TTTTCAGAAA ATAAGCAGGA TTTATTCCCT GAAATTACTG	3060
AATTAATCA CTGTTTCGAT TACTTTTGC AATATTAAAA GTAAATATTT AAACAGGTAA	3120
AAACAGAAAT AATGGTAGGG TCCTTATCAT CACCGTGAAT TCCAAGCTAG CATAGACACT	3180
AAACCTAGAG ATTCACACTA GAATGAAAGC TGGGAGAGCA GAGGAGTCTC AGAAGGATGT	3240
GGAGGCCAAT GGACACCTGC AACCTCTCCA ACGAAATGCC TACCTCCTCT CACTGCAGCA	3300
TCCATCTCTG AGCCTCTCG CAGCAGAGCT ATAAATTAG CCTGGCTCCT CCGTTCCAC	3360
ACATCCACTC CTGCTCTCCC TCCTCTCCTC CAGGTGACTA CAGTTATGAG GACCCTCAC	3420
CTCCTCTCTG CCTTTCTCCT GGTGGCCCTT CAGGCCTGGG CAGAGCCGCT CCAGGCAAGA	3480
GCTCATGAGA TGCCAGCCC GAAGCAGCCT CCAGCAGATG ACCAGGATGT GGTCAATTAC	3540
TTTCAGGAG ATGACAGCTG CTCTCTTCAG GTTCCAGGTG AGAGATGCCA GCATGCAGAG	3600
CTACAGACTA GACAGAAGGA CAGGAGACAG GCTCTGGAAT TGGATCTCAG TGGCAGATGT	3660
CACTTAGGTG GCTATACTTA ACATCTCTGG TCCTGGATTT TCTCATATCT AAATGGAATA	3720
GAGAACAAA GAAATCTAAG AGATTTTCT TTCTCCAAAA ACTTGATTCC AAGATATGAC	3780
TGTGAAATTC ACTAGATTAA AGATATAAGG AGATGCTACC TAGTTCCCTC TGGAGCCAGA	3840
CAAACAAGCT TAAGTATATA GGAAAATATT TCACCCGTGC TATATAGGAG GTTTAGAAC	3900
CTGGAGAGGA GCCTAAGAAT GTGTTCAGGT GTGTGTGTGA TGGGCAGGAA TGCAGAAAAG	3960
TGAAGCAAAG GAGAATGAGT CTCGAATCCT GTGTGACCAG CACTGCTCTG TGTATTTATT	4020
CCTATTGACT GAGATTGTT GTGCTACCGG CTGTAATACA GCCAACATCA CTCATCAGCC	4080
AACATGTGAC TTCTCCAAGA TTCCCTTAC CACCCACTGC TGNACCCCGT ACTCAGTTTC	4140
TGATGCTCTC TCTGGTCCC CAGGCTAAC AAAGGGCTTG ATCTGCCATT GCAGAGTACT	4200
ATACTGCATT TTTGGAGAAC ATCTTGGTGG GACCTGCTTC ATCCTTGGTG AACGCTACCC	4260
AATCTGCTGC TACTAAGCTT GCAGACTAGA GAAAAAGAGT TCATAATTTC CTTTGAGCAT	4320
TAAAGGGAAT TGTTATTCTT ATACCTTGTGC CTCGATTTC TGTCTCATC CCAAATAAAT	4380
ACTTGGTAAC ATGATTTCCG GGTTTTTTT TTTT	4415

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTIC:
- (A) LENGTH: 453 BASE PAIRS
 - (B) TYPE: NUCLEOTIDE
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: cDNA

- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CTCTGCCAAC TCTGGTAGCC TCACGTAGCT TAACAATCTG TGACTACAGT T ATG AGG Met Arg 1	57
ACC CTC ACC CTC CTC TCT GCC TTT CTC CTG GTG GCC CTT CAG GCC TGG Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln Ala Trp 5 10 15	105
GCA GAG CCG CTC CAG GCA AGA GCT CAT GAG ATG CCA GCC CAG AAG CAG Ala Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln Lys Gln 20 25 30	153
CCT CCA GCA GAT GAC CAG GAT GTG GTC ATT TAC TTT TCA GGA GAT GAC Pro Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly Asp Asp 35 40 45 50	201
AGC TGC TCT CTT CAG GTT CCA GGC TCA ACA AAG GGC TTG ATC TGC CAT Ser Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu Ile Cys His 55 60 65	249
TGC AGA GTA CTA TAC TGC ATT TTT GGA GAA CAT CTT GGT GGG ACC TGC Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly Thr Cys 70 75 80	297
TTC ATC CTT GGT GAA CGC TAC CCA ATC TGC TGC TAC TAA GCTTGCAGAC Phe Ile Leu Gly Arg Tyr Pro Ile Cys Cys Tyr * 85 90 95	346
TAGAGAAAAA GAGTTCATAA TTTTCTTGA GCATTAAGG GAATTGTTAT TCTTATACCT	406
TGTCTCGAT TTCCTGTCCT CATCCCAAAT AAATACTTGG TAACATG	453

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 94 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (ix) FEATURE:
 - (A) NAME/KEY: SIGNAL PEPTIDE
 - (B) LOCATION: 1..19
- (ix) FEATURE:
 - (A) NAME/KEY: PRO REGION
 - (B) LOCATION: 20..63
- (ix) FEATURE:
 - (A) NAME/KEY: MATURE PEPTIDE

(B) LOCATION: 64..94

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Arg Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln
1 5 10 15

Ala Trp Ala Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln
20 25 30

Lys Gln Pro Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly
35 40 45

Asp Asp Ser Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu Ile
50 55 60

Cys His Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly Gly
65 70 75 80

Thr Cys Phe Ile Leu Gly Arg Tyr Pro Ile Cys Cys Tyr
85 90

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: SIGNAL PEPTIDE

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Arg Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln
1 5 10 15

Ala Trp Ala

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PRO REGION

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln Lys Gln Pro
1 5 10 15
Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly Asp Asp Ser
20 25 30
Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu
35 40

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 AMINO ACIDS
(B) TYPE: AMINO ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: MATURE PEPTIDE

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Ile Cys His Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly
1 5 10 15
Gly Thr Cys Phe Ile Leu Gly Glu Arg Tyr Pro Ile Cys Cys Tyr
20 25 30